COMMUNITY HEALTH SURVEYS
A Practical Guide for Health Workers

2. Survey Sampling

SECOND EDITION

Previous Title:

SAMPLING: HOW TO SELECT PEOPLE, HOUSEHOLDS, PLACES TO STUDY COMMUNITY HEALTH

A Guide for Health Workers

Prepared for the International Epidemiological Association

by

W. LUTZ

Director, Medical Computing and Statistics Unit

University of Edinburgh Medical School, Edinburgh, Scotland

in collaboration with the World Health Organization

Published by

INTERNATIONAL EPIDEMIOLOGICAL ASSOCIATION

IEA

1982

Request for copies of this publication can be made to one of the Regional Offices of the World Health Organization listed below.

Regional Office for Africa, P.O. Box No. 6, **Brazzaville,** Congo.

Regional Office for the Americas/Pan American Sanitary Bureau, 525, 23rd Street, N.W. **Washington,** D.C. 20037, USA.

Regional Office for the Eastern Mediterranean, P.O. Box 1517, **Alexandria** - 21511, Egypt.

Regional Office for Europe
8, Scherfigsvej, 2100 **CopenhagenØ** - Denmark.

Regional Office for South-East Asia, World Health House
Indraprastha Estate, Mahatma Gandhi Road, **New Dehli**-110002, India.

Regional Office for the Western Pacific, P.O. Box 2932, **Manila** - 2801, Phillippines.

Printed in Switzerland. Second edition 1986.

COMMUNITY HEALTH SURVEYS
A Practical Guide for Health Workers

Other numbers in this series :

Number 1. Planning and Organizing

Number 3. Using Available Information

Number 4. Questionnaire Design

Number 5. Interviewing and Recording

Number 6. Presenting Survey Information.

SURVEY SAMPLING

INTRODUCTORY REMARKS :

Handbooks in this series have been initiated to encourage the health worker to study his community and to guide him in organizing a health survey to obtain the necessary information. They were started in order to provide him with the knowledge required to do this correctly, however, in a simple and inexpensive way. "Knowing my community well" is starting point in planning community activities towards reaching "Health for all by 2000". Indeed, a well-organized community survey should become an essential element of any country's H/2000 strategy.

There are right and wrong ways of collecting the information. There are right and wrong ways of interpreting and using it in management health of a patient, family or community. It is often the case that information collected is not even utilized. This refers also to information collected by undertaking a health survey.

Before we collect information it should be agreed as to how and for what purpose the information will be used.

You will see that the survey workshop organized by Dr. Ashma (page 17) also begins with discussion on using information. Sampling and sample size follow as the second and third sessions.

In this series, however, the sampling booklet appears **second** and not third. The reason is simple. There are right and wrong ways of taking a sample. We have been urged repeatedly by health workers to tell them how to do it correctly, hence the reason for hurrying the printing of this book. There is a slight inconvenience in that a few paragraphs, which explain further reasons for Dr. Ashma's workshop, appear in the "Using information" booklet, which was originally planned to appear before the "Sampling" ones. We apologise for this.

Dr. B. Skrinjar-Nerima

Workshop on:

HOW TO SELECT PEOPLE, HOUSEHOLDS, PLACES TO STUDY COMMUNITY HEALTH.

INTRODUCTION

A survey, sometimes called a sample survey, is undertaken to get information about people, their health and about the way they live and work. Usually it is not possible to visit or examine everyone because it would take too long, would cost too much or because there are not enough assistants to do this. Instead a **sample** *of persons or homes is selected and only these are examined or selected.*

This sample will give reliable information provided that:

(1) the sample is representative, i.e. reflects correctly all sections and kind of persons belonging to the group, or community being studied:

(2) the people included (selected) in the sample are approached in the correct manner so that they will help and co-operate with the investigators (doctors, nurses or interviewers);

(3) the sample is large enough.

Simple methods are known that will increase peoples' co-operation and interest in a survey and there are ways of asking questions that improve the completeness and correctness of the information given. These methods are discussed in two other booklets of this series:

(i) Designing Questionnaires

(ii) Interviewing and Recording

This booklet is mainly concerned with three selection (sampling) problems and these are:

(i) Describing different methods of obtaining a representative sample, sometimes called **random** *sampling methods.*

(ii) How to decide which of these different sampling methods to use.

(iii) Determining how large the sample should be, i.e. how many persons, families or houses should be seen and examined.

The layout of this book is similar to that followed in "Planning and Organising a Health Survey". General principles and important questions are set out on the left hand page. The right hand page describes practical sampling problems and surveys in which these general principles are applied.

Readers may find it easier to first read the left side pages of any section and then to read the application described on the pages of the right side.

Notes

GUIDING PRINCIPLES

The Need to Discuss and Compare Surveys

Survey theory provides guidelines on how to plan and organise a survey, how to determine the sample size, how to design the questionnaires and how to extract the information and report the results.

Surveys benefit greatly from practice and experience. For this reason survey planners and organisers should, whenever possible, meet with other persons doing surveys so that they can discuss their methods and compare their results. It is also important to learn from others why they carried out their survey in a particular way rather than some other way. At the end of a survey, organisers often feel they could improve their methods and that they would now do some aspects of their study differently. A great deal can be learned from discussions with other survey workers.

A survey "Workshop" is an excellent way of bringing people with similar interests together for discussion and learning. A workshop is a one to five day meeting or small conference at which invited speakers describe the work they have done, the problems they met and how they dealt with them.

EXAMPLES AND APPLICATION

Organising a Workshop

Because of the interest shown and because he wanted to learn from the experience of others, Dr Ashma decided to approach his government for support to organise a survey Workshop. This he obtained after explaining to the Minister of Health the importance of a workshop, how he intended to organise it and whom he was inviting to the workshop.

Dr Ashma invited six principal speakers. Five of these had carried out a health or social survey during the last two or three years. The sixth speaker, Dr Retlaw, was a survey specialist.

Invitations were sent to 60 persons known to be interested in health surveys and 27 were able to attend. The workshop was held in the city's Medical School which provided a suitable room with a blackboard and a projector for the speakers to illustrate their talks.

The workshop was to last four days; each day was organised into morning and afternoon sessions. At each session, a different survey topic was discussed. At the end of each session Dr. Retlaw was to summarise the most important points raised by the speakers and to present a simple summary of the main survey principles.

An important feature of a workshop is that ample time is set aside for general discussion and for questions to be put by the audience.

Frequently, workshops are divided into "morning" and "afternoon" sessions, each of which is devoted to some particular topic or problem.

A short workshop program leaflet was prepared and sent to everyone invited; this listed the main activities as follows:

Survey Workshop

Programme

Sessions		Topic for Discussion
Monday	Morning	*Using available information*
	Afternoon	*Sampling*
Tuesday	Morning	*The Sample Size*
	Afternoon	*Questionnaire Design*
Wednesday	Morning	*Interviewer Training*
	Afternoon	*Information Extraction and Survey Analysis*
Thursday	Morning	*Presenting and Reporting Survey Results*

This booklet provides for information on two sessions: Survey Sampling and Sample Size.

Notes

SAMPLING : GENERAL PRINCIPLES

THE IMPORTANCE OF CORRECT SAMPLING METHODS

The aim of every health survey is to obtain information about some health need or medical problem that is of importance to the community. For a survey about health, diet or living conditions it is necessary to select a sample (group) of people to interview (speak to and ask questions), or of patients to examine and study, or of homes to visit. For other kinds of surveys, e.g. surveys concerned with agriculture and crops, it may be necessary to select a sample of fields or of herds of cattle so that these can be examined for disease, for their yield or any other information.

The need for a representative sample to overcome selective reporting of illness.

Correct sampling methods are essential for a good survey

Correct sampling procedures ensure representative samles, i.e. ensure samples consisting of persons and families very similar to those found in the general community.

Incorrect method of sampling

There are right and wrong ways of taking a sample. The incorrect way is to allow the survey workers of the organiser to individually select, or to influence the selection of, the survey sample. The reason for this is that, for each of us, individual selection is influenced by personal convenience,

WORKSHOP ON HEALTH SURVEY

Session 2: Sampling

First Speaker: Dr. J. Ashma

Subject: The Importance of obtaining a Representative Sample

Dr. Ashma started by saying that some of his colleagues were not convinced of the need to obtain a fully representative sample when doing a survey. They felt that too much effort is required to ensure representative samples.

A truly representative sample is however essential. Dr. Ashma illustrated this by describing three studies in which the results would be distorted without a proper and representative sample.

Knowing his colleagues interest in whooping cough, Dr. Ashma used this as his first example.

For instance, to study the effect of whooping cough, it is *not* enough to examine only children brought to a primary health care clinic or a hospital. The reason for this is that only a section of the community or special types of cases report at a medical care centre. Very ill patients and those who believe the clinic can cure them are more likely to attend. The milder cases or children from more distant villages are less likely to visit the medical care centre. If the survey is to determine how many are affected by whooping cough as well as determine its natural history then the study must report on all cases, both the serious as well as the mild.

Dr. Ashma next described how misleading information will be obtained in studies of child nutrition and development unless great care is taken to get a representative sample of families from all sections of the community. The selection

opinions and preferences. These nearly always lead to an unrepresentative sample that does not accurately reflect the whole situation.

The need for a representative sample to overcome differences in wealth, social class and rural/urban characteristics.

The Importance of a Representative Sample

The purpose of a selection (sampling) procedure is to obtain a representative sample of the persons or places with which the study is concerned. Sometimes it is difficult to get a representative sample but it must be done.

The need for a representative sample to take account of within family differences as well as differences between families.

must ensure that the concentration (proportion) of each type of family in the sample, of rich and poor and of large and small, of rural and urban families must be very close to that in the community. As a further example, suppose a survey studied the way people live and work in a community of 110 families; 80 of these families live and work on the land whereas the other 30 live and work in the village. If the survey was only able to study 20 of these families then this sample of 20 families should contain approximately 16 families working on the land. The exact proportion is not always possible but the sample proportions must be close to the community proportions.

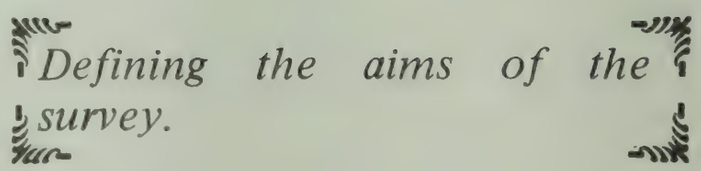
Dr. Ashma went on to say that some might argue that it is safer, and not much more difficult, to visit all 110 families and not just 20. This may be possible in this example if there is enough time and staff. But the situation is quite different if the doctor wants to do the same study in a community of 1 000 or 10 000 families. The expense involved and the number of staff required makes a complete study impossible; a sample survey is the only way.

Another example mentioned by Dr. Ashma was a study of child development up to puberty. It will be misleading (wrong) to seek out large families with many children in the hope that in this way there will be enough children to study but only a few families to deal with. This selection procedure is wrong because conditions affecting child nutrition and growth often differ between large and small families. The amount of food available, the kind of food prepared and general living conditions are often different in larger families compared to smaller families. The correct method is to use a selection (sampling) procedure that will give the same proportion of large and small families, of poor and rich families in the sample as are living in the community as a whole.

Notes

Define the Study Unit

Every health survey is concerned with specific groups of persons or 'study units' as well as a definite area or group of institutions.

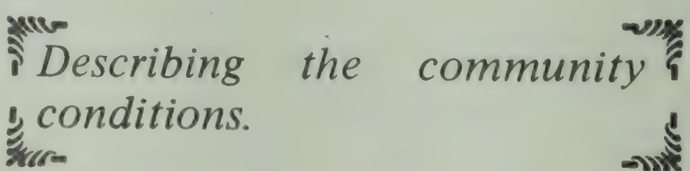

Defining the aims of the survey.

The first step in devising (making) any sampling scheme is to define the study unit.

This is done by:

(a) providing rules by which the persons, patients, families, farms, or any other 'study unit' being surveyed, can be recognised (identified);

(b) listing or describing the institutions, communities or areas to which the survey is confined (restricted).


Describing the community conditions.

All the study units taken together are called the 'survey population'.

The purpose of describing in detail, i.e. defining, the study unit is to:

(i) enable everyone concerned with the survey to recognise the kind of person, patient or 'unit' being studied;

Second Speaker: Dr. K. Singh

Subject: A Survey of School Performance

Dr. Singh's survey aimed to test performance in reading, writing and arithmetic in school children who had attended school for between two and three years. The survey was also to ascertain (find out) how many children obtained help with their school work at home and what the effect of this help was on school performance. The question had been asked whether girls got less help and encouragement at home than did boys, and if this was so, how did this affect school performance.

In this small town community, school attendance was not compulsory and both girls and boys could attend any one of seven schools. Some children attended for the first time when they were 9 or even 10 although most came when 6 to 7 years old. It was known that many children, perhaps 1/5 to 1/3 did not attend. Also poorer families and families who still worked some land outside the town were likely to withdraw their children from school for a few weeks whenever they needed their help.

Comparison of school performance would become very difficult if the children studied were not of similar age and similar school experience. The study unit was therefore defined as:

(a) A boy or girl who had attended school between two and three years.

(ii) make clear which kind of persons, patients or families are **not** *part of the study;*

(iii) make it possible to provide rules by which study units are selected for inclusion in the survey sample.

Defining the study unit:

1) Both sexes included
2) Age limits set
3) Criteria for exclusion/inclusion in the study

The survey planner must, at the outset of the study, make very clear which persons and patients are to be studied *and which are to be excluded from his survey. It is helpful to write down a short check-list such as:*

(i) Are both male and female to be included in the study?

(ii) Is the survey concerned with a specific age group?

(iii) Which communities villages or areas are to be covered by the survey?

List Sampling:

Have you checked that no names have been left out?

Sample Size:

How is the size decided? Are 300 children too much **or** *not enough?*

(iv) What are the diseases and disabilities of primary interests? Have guidelines on how to recognise these been given to junior doctors and nurses?

(b) During the 2 to 3 years at school the child had not missed more than 5 school weeks each year.

(c) The child at the time of the survey was between 8 and 10 years old

(d) The child was generally in good health and was not mentally or otherwise handicapped.

Dr.Singh went on to explain how he drew up a sampling list for his survey.

One of his survey assistants spent a day at each of the town's 7 schools and with the help of a teacher made a list of names of children who satisfied the definition (criteria or conditions) of the study unit as described above. For reasons that will be explained later a separate list was made for boys and for girls, i.e. 2 lists.

Dr Singh felt that for practical reasons he could not deal with more than about 300 children. The three main reasons for this were:

(i) The willingness of teachers to give a special reading, writing and arithmetic test to the survey children.

(ii) That it was important not to disrupt school routine too much

(iii) That there were only three survey assistants who could question the children about the amount of help they got at home.

(v) If the patient also suffers from certain other diseases and disabilities, should this person be excluded from the study? Have guidelines been set out explaining whom to exclude?

(vi) Is the severity or duration of the disease and disability of importance to the study? Are chronic or recurrent conditions to be included or excluded?

After describing the study unit and the institutions and communities to be included the survey planner must then consider ways (methods) of getting his sample. There are many sampling procedures but a few simple methods are enough to deal with most local survey situations.

The simplest of the survey sampling methods is **list sampling.** *This is often effective for studies done in schools, hospitals and clinics.*

Obtaining a representative sample by drawing numbers from a box.

List Sampling

The sampling list is obtained by constructing (making) a list of all the individuals or study units belonging to the survey population. From this list **select a number of persons (or study units) at random;** *these will be examined or interviewed as required by the survey. Sometimes a list is already available and can be used by the survey organisers.*

'Select at random' means selection by a method based on chance. *Because selection is by chance, i.e. at random, neither the planner or his assistants or the interviewers can influence who is included in the survey sample and who is left*

Typical problem

What to do with 'not available' or absent cases.

Actual sample size not equal to the planned sample size.

A sample of 150 boys and also a sample of 150 girls was drawn.

Dr Singh went on to describe how he actually drew the sample. He was very concerned that no one should influence which boys and girls were tested and interviewed. The list of boys, suitable for inclusion in the survey, contained 816 names and the list of girls had 471 names. Dr Singh knew fewer girls attended school but he was surprised the difference was as great as these lists showed.

He first wrote out the numbers 1 to 471, each number on a separate piece of paper, and folded these so that the number could not be seen before placing them in a box and mixing the pieces of folded paper several times. He then drew 150 of these folded papers from the box. The girls on his list with these numbers were then included in the study.

A similar method was used for drawing a sample of 150 boys.

The interviewers spent a day at each of the 7 schools. At each school the sample of boys and girls were interviewed individually as well as given a short reading test. Later they were brought into a classroom and given a short writing and arithmetic test.

On the days the schools were visited there were always a few boys and girls absent who should have been included in the sample. Dr Singh left them out of his survey and he did *not* replace them with other boys and girls.

At the end of the survey, he had a sample of 138 boys and 133 girls.

out. In this way every aspect under investigation has its proper opportunity (chance) to be seen and examined. The private opinions, attitudes, interests and convenience of the survey organisers and interviewers is in this way prevented from interfering with the selection of a representative sample.

The easiest way of selecting a random sample of names from a list is:

(1) To number in sequence all the names on the list, starting with 1, 2, 3,... and so on until a different number has been placed against each and every name on the list.

(2) To select the sample from this list.

This can be done in one of two simple ways which are described in the Appendix.

Keeping Lists Up-to-date

List sampling is easy and effective provided a complete list of sampling units is available or can be drawn up (made) by the survey organisers. However drawing up a list can be expensive and can take a long time. For this reason it is an advantage to use existing lists if these can be found. **Lists very soon get out of date.** *People change their address and change their work; family composition changes by marriage, births and deaths. For many reasons survey populations change and the list for the survey population must change also.*

Official lists from cities and other authorities are often incomplete, i.e. do not list every study unit in the survey population. All lists should be checked before using them to make sure they are good enough for the survey.

Arrangements must be made for updating lists, *i.e. to add any new names or take out old names when persons move away or for some reason no longer belong to the survey population. This is particularly important for studies that continue for a long time, e.g. a year or more.*

Dr Singh went on to emphasise that he had deliberately chosen his sample of girls to be as large as his sample of boys; he had planned for 150 of each. He wanted the same amount of reliable information on each because the survey aimed to compare female with male school performance. This had been achieved by the sampling scheme. If he had taken a simple sampling rule such as choosing 1 in 4 of the boys and girls on his list (i.e. a fixed sampling fraction) he would have unequally sized samples of about 205 boys and 120 girls. For comparing groups it was better to have approximately the same amount of information on each.

Notes

SAMPLING BY 'NUMBERED TAG'

A frequent problem in medical surveys is the selection of a representative sample of patients attending at a hospital or clinic. For instance, lists of bed occupancy are usually available and these can be sampled similarly to the method described in the previous example. However, at a busy out-patient clinic, patients may arrive without appointment at any time during the day; they sometimes arrive without a previous clinic record. Situations such as this require a different sampling method which does not need a list of patients.

Describing the survey aims and purpose

Sampling by numbered tags is simple and it is an effective way of sampling out-patients in hospitals or clinics. The 'tags' are small discs or flat squares of wood, metal or other suitable material and on each tag a number is printed. The tag numbers start at 1, 2, 3, and so on up to any number required, however big.

In the clinic situation numbered tag sampling proceeds as follows:

Step 1

A tag is given to each patient as they arrive at the clinic. The first patient gets the tag numbered 1, the second gets the tag numbered 2, and so on. A tag will be issued to each patient no matter at what time they arrive. At this stage it is not necessary to check whether the patient belongs to the survey population or not; everyone arriving for treatment at the clinic is given a numbered tag.

The exact number chosen at each clinic session is not known beforehand

Third Speaker: Dr. A. Xuma

Subject: A Survey of Parasitic Disease Amongst Children Attending an Infant Care Clinic

Dr. Xuma started his talk by explaining that a young doctor had recently joined the Health Centre. After about six months he thought more than three quarters of young children under 3 years seen at the Infant Care Clinic were infected with gut parasites of one kind or another. Dr. Xuma knew parasitic infestation to be high but was unwilling to accept that it affected more than 10% to 15% of children in this age group.

There were several difficulties. Although the Clinic urged (encouraged) all mothers to bring their babies for regular two monthly check-ups, only about 10% did so unless the child was definitely ill. The children seen at the Clinic were therefore not typical of infants in the community. Also, testing for gut parasites required a stool test. Laboratory facilities for doing this were scarce; a study must not overload the services or delay specimen testing for other patients.

Another difficulty was that mothers could not be asked to bring back a stool specimen the next day; few would be able to come back such a long way. Mothers would have to remain at the Clinic after the doctor had seen each child and wait until the infant defecated. This might hold up the flow of patients as waiting space and other facilities at the Clinic were very overcrowded. For these reasons, only a small sample of infants could be taken at each Clinic session. To provide a sufficient sample size the survey was spread over 4 or 5 weeks. A long survey period meant that only a few mothers were required to wait at each Clinic session and the laboratory only needed to do between 4 and 9 extra stool tests each day.

Step 2

The survey organiser decides on the **sampling fraction,** *i.e. on the proportion of patients to be taken into the survey sample for further interviewing and special examination. This is usually 1 in 10, or 1 in 5, or some other suitable proportion (fraction).*

For example, selecting each patient whose tag number ends in 7 results in a 1 in 10 sample. Likewise any other single digit between 0 and 9 will result in a 1 in 10 sample. Similarly, a '1 in 3' sample fraction is obtained approximately by choosing any **three** *figures (digits) between 0, 1, 2, ... 9. Patients whose tag numbers end in any one of these three figures are then included in the sample, the others are not.*

Defining the Study Unit:

Inclusions:

1. *Both sexes*
2. *Age limits*

Exclusions:

1. *Had previous visits during survey period*
2. *Has severe abdominal disorder*

The Sample Size

Discrepancy between actual and planned sample size

Step 3

Each patient hands in his tag as he is seen by the doctor. The doctor will provide treatment in the usual manner but will in addition determine (find out) whether:

The 'Numbered Tag' sampling procedure applied

Defining the study unit for this survey was easy:

(a) all infants attending the Clinic and not suffering from a severe abdominal disorder;

(b) infants aged between 6 months and three years;

(c) this was the first time during the survey period that the child had been brought to the Clinic.

Between 120 and 180 infants and young children were brought to the Clinic each day. Of these about 50 to 70 would be aged 6 months to 3 years although it could be as few as 30 or as high as 90. A sample of one in ten of infants in the right age group would select between 5 and 7 cases for inclusion in the study at most sessions. This would not strain the Health Centre services. The survey was to continue for 30 Clinic sessions and was expected to bring about 180 infants into the parasite study.

At the end of the survey it was found that a 1 in 10 sampling procedure selected 193 infants for the study. However, three stool tests failed in the laboratory and one mother left the Clinic without handing in a stool specimen, so the actual sample size at the end of the survey was 189 infants tested for gut parasites.

As each child was seen for diagnosis and treatment, the doctor at the same time noted whether the patient belonged to the survey population, i.e. satisfied the definition (conditions) for a study unit. If the child belonged to the survey population then it was included in the sample if its tag number ended in '3', e.g. 3, 13, 23, 33 and so on.

(i) *this patient is part of the survey population;*

Changing the selection tag numbers so that supporting staff do not know who will be included in the study

(ii) *the patient's tag number ends in one of the figures that selects patients for inclusion in the sample;*

(iii) *the patient has not already been attended to or been included in the survey sample at a previous clinic session* **during the survey period.**

Patients who meet all three conditions are part of the survey sample and are included in the survey for special interviewing and examination.

Step 4

At the end of the clinic session all the tags are collected and put in order 1, 2, 3, ... and so on, ready for use again at the next clinic session. **It is important that the selection numbers be changed at each session.** *The person handing out tags to the clinic patients should not know what the selection numbers are. If they did know then they could select which patients got tags ending in these numbers and so decide who came into the survey sample.*

Repeat Visits by Patients

In hospital and clinic studies it is important not to include the same patients in the survey sample more than once. A patient should not have more than one chance of being selected. *This could easily happen in sessions. This is particularly true of the chronically ill and of the more severe cases who attend more frequently. Patients have a chance of being selected each time they attend unless arrangements are made to prevent this. To stop this happening,* **patients may only be selected the first time they attend the clinic during the survey period.**

At their second or repeat attendance during the survey period they are of course treated but are no longer chosen for the survey.

This procedure was repeated at each Clinic session except that each day the number chosen to select the sample was changed. The first day it was '3', the next day it was '7' and the following day '5'. The day after that the doctor decided in future to draw a new random number between 0, 1, ..., 9 before each session started; this number would be used for one session in the same way '3', '7' and '5' had been used during the first three days of the survey. After a few sessions some of the numbers drawn would be the same as those already used, but this did not matter as long as none of the assisting staff knew what the number was.

The results of the survey showed that of the 189 stools tested, 38 showed positive results for parasites, i.e. about 20%. This was higher than Dr.Xuma had thought but was very much less than the figure (75%) claimed by the young doctor. Nevertheless 20% was disturbing and it was agreed to do a community survey the following summer when parasite infection was likely to be at its highest.

Notes

STRATIFIED SAMPLING

Stratified sampling is another sampling method. It is especially useful when similar institutions or organisations are to be studied and particularly if comparisons are to be made between them.

Strata useful for comparison studies

The word **strata** *when used in connection with sampling means a group of study units or of people who are rather similar or who are working or living under similar conditions. In stratified sampling the population is divided into* **groups** *of study units that are rather similar or are placed in similar situations. The strata may be very dissimilar (different) from each other but within each strata the study units are exposed to similar conditions or are more alike in some respect. After the whole population has been divided into two or more strata, a random selection of study units is taken from* **each** *strata* and a similar survey is carried out in each of the strata.*

Each Health Centre is considered as a separate strata for comparing area health needs

The basic steps for stratified sampling are:

1. *Divide the population to be surveyed into strata of similar study units or into areas within which similar social, environmental or health conditions exist. Often these*

The sampling fraction is 1 in 20. No exclusion criteria in this survey

Separate surveys in each strata

* *Strictly speaking 'stratum'. For simplicity we use the plural 'strata' even when only a single stratum is meant.*

Fourth Speaker: Dr. J. Desai

Subject: Sampling Procedures used in two Clinical Surveys

Dr. Desai explained that he was a teacher in a department of Public Health and as such was often asked for his advice on conducting health surveys. He particularly wanted to illustrate the usefulness of using a stratified sampling plan for comparing clinical and general health problems presenting at clinics in different areas. The same methods could also be used to compare the experience of general practitioners (doctors) working in different parts of a large city.

Dr. Desai went on to describe a survey designed to compare the health problems at three Health Centres. These Health Centres served different communities, one along a large river and well irrigated area, one serving a mainly agricultural community in a relatively dry region, whilst the third Health Centre served a coastal town and its environment whose population mainly worked as fishermen or in the coconut industry. A separate survey was made of patients reporting at each of these Centres; the 'numbered tag' sampling method was used. The survey was done over a period of one year and a one in twenty sampling fraction was used. All patients reporting were eligible for inclusion in the study. During this time an identical questionnaire to record a standard medical history of patients was used at each of the three Centres.

This resulted in three separate surveys, one at each Centre; these were easily comparable because the same questionnaire was used and similar examinations done at these Centres.

strata are institutions or organisations and the groups of people within them.

Each General Practice is considered as a separate strata

2. *Either make a separate and complete list of the study units within* **each** *strata and from* **each** *strata draw a* **separate** *random sample of study units using these lists*

   OR

   Use the numbered tag method to draw a separate sample of study units from **each** *strata.*

3. *A similar survey is then done on the sample of study units in each of the strata, i.e. the same questionnaires are used, patients are examined in the same way and subjected to the same tests.* **In effect, a similar survey is carried out in each of the separate strata.**

How many strata should there be ?

Stratified sampling has two main advantages:

Any number of strata are possible

1. *More reliable information is obtained for the same sample size if the survey population can be divided into easily recognised strata within which the study units are more alike than they are for the population as a whole.*

2. **Comparisons between strata are easy.** *This is so because:*

There are advantages in only using a few strata

As a further example Dr. Desai mentioned that general practitioners, especially in some larger cities, often discuss the possibility of doing some joint research. It is often possible for general practitioners to get a list of their patients and to agree to take a sample of cases from this list for a special study. Alternatively they can use the numbered tag method to select patients coming to their surgery over an agreed period, perhaps several months. Provided the general practitioners all use similar questionnaires and agree to do the same clinical investigations, it is then possible to compare their practice experience. It is also possible to combine their separate results to provide a city survey of ill-health reporting at general practitioners surgeries. Of course the doctors might have specialised interests and agree to concentrate their study on particular kinds of disease or special groups of patients.

After Dr. Desai finished, a questioner from the audience asked what is the best number of strata to form. Dr. Desai answered that it depended upon the survey being done.

For instance, in the survey of school performance described earlier in the session by Dr. Singh, only two strata had been formed, one list for boys and one list for girls. In his own Health Centre survey, three strata were formed whereas in the general practitioner study, seven practices participated, i.e. there were 7 strata.

As a general rule it was best to restrict the number of strata to the number of comparisons to be made. More strata require more survey supervision and organisation. At a later stage too many strata lead to more difficult calculations when analysing the survey results.

(a) a separate but similar survey is done in each strata;

Sample size

(b) the sample size taken from each strata can be adjusted to suit the aims of the survey. A relatively larger sample may be taken from very important strata or from very small strata so as to ensure reliable comparisons.

An unsuitable sampling fraction can lead to unnecessarily large surveys

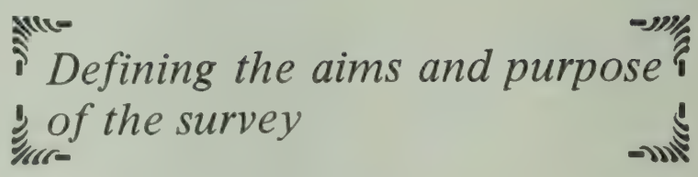
No Lists Available

For many surveys no suitable list exists and numbered tag sampling is inappropriate. It may also be difficult and expensive for the organisers to make their own lists; sometimes it is impossible. In these circumstances 'cluster sampling' can be very useful.

Another questioner wanted to know why the sampling fraction had been set at 'one in twenty'. Dr. Desai explained that at two of the Health Centres between 200 to 250 patients reported each day. The third Centre served a smaller population and between 100 and 150 cases presented daily. Most of these cases were of course treated as out-patients. Selecting 1 in 20 gave a sample of between 10 and 13 survey cases on most days at the larger Centres and between 5 and 8 at the smaller Health Centre. Over a year this provided a sample of 4113 and 3602 at the two larger Centres and 2130 at the smallest. He now felt the sample was larger than was needed for making comparison between the Centres. At the two larger Centres a sampling fraction of 1 in 40 would have been sufficient.

Notes

A **cluster** *is a group of study units close together, i.e. crowding together in the same area or neighbourhood. As an example, a village is a cluster of families because all these families are living in the same area.* **Cluster Sampling** *is always done in three stages.*



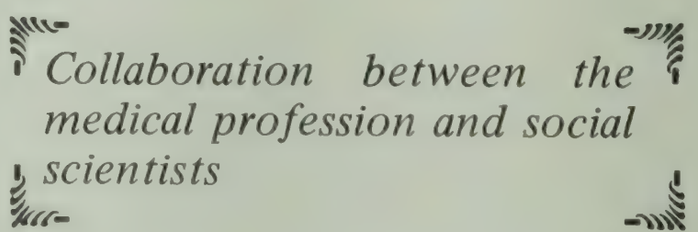
(i) divide the survey area, such as district, several villages or a town, into clusters. Number these on a map or make a list of these.

(ii) draw a random sample of clusters from these.

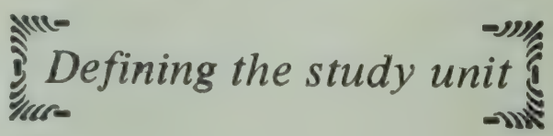
(iii) arrange for the field workers, usually nurses, doctors or interviewers, to visit **each and every one** *of the study units within the selected (sample) clusters.*

Cluster Size

The size of a cluster is the number of study units within it.



As a general rule, the cluster size should be small. *It is far better for the survey results and for managing the field work, for clusters to contain fewer study units. This means forming (making) many clusters in the survey area with each cluster containing only a small number of study units.*



Fifth Speaker: Dr. T. Tamburi

Subject: A Survey on Urban Housing and Living Conditions in relation to T.B. and Eye Infection

Dr. Tamburi explained that the Social Science Department of his University was interested in studying the housing conditions in a town of about 40,000 inhabitants. It was known that severe poverty existed. The Social Science Department was particularly interested in:

(a) The number of persons resident (eating and sleeping) in each separate unit of residence. A unit of residence could be a house, hut, room or apartment.

(b) The cooking facilities available such as a stove, kitchen sink, or tub; the water supply as well as refuse (rubbish) disposal and drainage.

(c) The toilet and washing facilities

(d) The number of seriously ill, housebound cripples or very elderly living in the town; they were also interested in the care these ill and disabled persons received.

Because of the health aspects of the study, Dr. Tamburi discussed the survey plans with the two town doctors. Both doctors became very interested and in particular asked that in return for their help the survey should also include an examination for T.B. and for eye infection. The doctors would need to visit all the survey families (households) and medically examine these individuals, provided they agreed. This would help the survey to ascertain ill-health and disability.

The study unit for this survey was a house, hut, room or apartment in which a family lived together with any relatives and lodgers; to 'live' in a place (address) meant these persons prepared their food, ate and slept at this

Large villages are often divided into two, three or more areas. Each of these smaller areas is also a 'cluster', just as the whole village is a cluster, only the cluster sizes are now much smaller.

List sampling not possible

People living within the same cluster are more likely to live and work under similar conditions than people living in clusters further away. To obtain a representative sample it is necessary to study all the different parts of the community and not see too many people living under similar conditions. Small clusters make it possible for the sample to contain more clusters spread throughout the survey area without having to investigate too many survey units.

Using a city map to draw cluster boundaries

Deciding on Cluster Boundaries

In a large district, clusters may be very different socially and geographically. A large district will contain several villages, the larger of which can also be divided into several clusters. All the clusters taken together must:

(i) **cover the whole of the survey area**: *no part must be left out;*

(ii) **not overlap**; *every study unit (e.g. family) is inside a cluster but no study unit belongs to more than one cluster.*

Before a list of clusters can be made it is necessary to decide on:

(i) **The appropriate size of cluster required.** *The typical cluster size chosen will determine how many clusters are listed.*

Typical Problem:

Lack of planning information

address. For a survey of this kind, it was impossible to construct a list of study units. The field workers, in this case a doctor or his trained nurse, together with a Social Science Department interviewer, would have to establish at the time of their visit, of what the 'home' or living place consisted; that is establish how many persons lived, ate and slept there. A cluster sampling method seemed the only possible way to carry out such a study.

Fortunately the City Authorities offered copies of a street map for sale. It was about 5 years old and did not show all the small side streets; in some of the empty spaces shown on the map homeless people had built themselves huts. In other areas the map still showed houses which had been torn down. Despite these faults, the map was valuable as it showed all the important streets, their names and also indicated the most important buildings and city landmarks.

Dr. Tamburi spent a few hours marking cluster boundaries on this map *in pencil*. He then sent out two of his staff with this map to tour (travel through) the town. They were to make cluster boundary changes on the map where it seemed people lived more or less densely (close together) than the 5-year old map indicated. This took three days. The cluster boundaries were then drawn in ink on the map and numbered in order. Someone else checked that every cluster had a number and no number was used twice. In all 432 clusters were formed and these covered every area in the town.

This still left Dr. Tamburi to decide how many clusters to take into his survey sample. He did not know the cluster sizes, i.e. how many families, their relatives and lodgers lived as separate households in an average cluster. Without this information it was not possible to plan the size of the survey.

The cluster size also determines how many study units will have to be examined or interviewed for every cluster drawn into the sample.

(ii) **Where to draw the boundaries of these clusters.** *The boundaries of every cluster must:*

(a) be easy to recognise by the field workers;

Pilot study to determine size and other cluster characteristics

(b) enclose about the same number of study units for each cluster. This is not always easy to do but clusters should be made as similar in size as possible.

Marking the Cluster Boundaries

In practice, it is necessary for a responsible (senior) person to travel through the survey district and to draw a simple map that shows how each village or area is to be divided into clusters. This map must show the boundaries of the clusters and also give the names of villages, streets, rivers and so on. The map should also indicate (show) other features that will help interviewers to recognise the cluster boundaries. These features may be a railway line or a school building, a well, some very large trees, a wall or anything that is easily recognised as part of a cluster boundary.

Preliminary pilot study estimates of cluster size, etc.

Uneven cluster sizes

To get this information he did a small pilot study, choosing four clusters at random out of the 432 clusters. He asked his two survey assistants to spend two days visiting these four clusters to obtain information on:

(a) The number of separate households in each of these four clusters.

(b) The number of persons living in each of the households.

(c) The number of households in which at least one person was suffering from T.B. or an eye infection.

The results of the pilot study were set out in a small table:

	Cluster Number on the map	Number of households in the cluster	Number of persons living within the cluster	Number of households with at least one infected person
1.	37	22	192	5
2.	141	13	142	2
3.	309	8	83	2
4.	402	12	131	1
		55	**548**	**10**

Dr. Tamburi was a little disappointed in these results. He had aimed to form clusters with about 10 households in each. Clearly some, like the first of these clusters, had turned out to be much larger. However he felt he really could not improve on this and decided to continue the survey despite the uneven cluster sizes. Although evenly sized clusters have certain advantages which lead to more efficient sampling, i.e. more information is collected for the same effort, it is not wrong to have clusters of uneven sizes.

Government and Local Authorithy maps should be used if they exist. *The cluster boundaries and other features can be marked (drawn) on the map. Check that the map is up to date and is not so old that it no longer shows all the important road, agricultural and town features.*

Local Support and Knowledge

People living and working in the survey district can be very helpful in deciding on cluster boundaries. Local people often have a good idea of how many families live in the various part of a village or town.

They can use this local knowledge to advise where to draw cluster boundaries so that the clusters will all be about the same size. Local people can also give the names of streets, buildings and other boundary features so that it is easier for interviewers to find places.

Sample size and estimate of survey field work

Numbering the Clusters

Before drawing the random sample of clusters, give each cluster a number, starting with 1, 2, 3 and so on in sequence until every cluster has a different number. This can be done in one of two easy ways.

(i) write out a list of clusters and then give a number to each entry on the list;

Or

(ii) number the clusters on the map showing the cluster boundaries.

Most survey organisers prefer the second method.

Typical Situation:

Preliminary pilot study results differ from the final survey findings

Typical Problem:

What to do with the pilot study results ?

From these results it seemed as if the typical cluster size was about 14; this value was obtained by dividing the total number of households by the number of clusters visited, i.e. 55/4 = 13.75. Fourteen was close enough to 13.75 for planning the study.

The pilot study estimate of the level of infection was 10/55 = 0.18, i.e. about 18% households seem to have at least one person with T.B. or a more serious eye infection.

The town doctors felt that if each of them had the help of a nurse and an interviewer from the Social Science Department they could between them manage to medically examine and interview six households a day. This meant they could cover about 300 households in the ten weeks planned for field work. The cluster sample size would therefore have to be 300/14 = 22 approximately. Because the pilot study was very small and because Dr. Tamburi wanted to be sure he had enough clusters to be representative of the whole town, he took a slightly bigger sample of 25 clusters. These were chosen from the town's 432 clusters by using a random number table. Dr. Tamburi allowed 13 weeks for the study.

The survey then proceeded as planned although the field work took 17 weeks instead of 13. In practice it was not always possible to complete 6 households each day; sometimes only 4 could be done. Also most cluster sizes turned out to be nearer 16 or 17 households and not 14 as the pilot study had indicated.

After Dr. Tamburi had finished, a questioner from the audience wanted to know what had been done with the results of the four pilot study clusters. Had they been included in the sample of 25, i.e. only 21 additional clusters selected? Or had they been included in the final results of the survey so that the survey information was finally based on 29 clusters?

Cluster Sample Sizes

How many clusters to take into the sample depends on how many study units are needed for the survey. For example, consider a survey in which the study unit is the family and 300 families are required. If the most usual cluster size is 20 families, then a sample of 15 clusters is needed. After selecting 15 clusters it will be found that there are not exactly 300 families in the study. This is because not all clusters contain exactly 20 families. If the clusters have been carefully drawn to be about of the same size, then the number of families should be close to the planned number. This emphasises the need to spend time and care in deciding on cluster boundaries.

Typical Problem:

Arrange to test the questionnaires during the pilot study

Selecting the Clusters

After deciding the cluster sample size, the random selection of clusters is done by list sampling.

Numbers are drawn from folded pieces of paper in a box or from random number tables (see Appendix); those clusters with the same numbers as the numbers drawn are included in the survey sample. All other clusters are excluded and are not studied during the survey.

Difficulties with Cluster Sampling

Cluster sampling is a very effective and easy way to carry out a survey provided it is possible to:

(i) draw or get a map of the area;

(ii) divide the area into clusters of similar size with easily recognisable boundaries;

(iii) draw small clusters and so prevent the inclusion of too many study units in the survey.

Dr. Tamburi replied that neither had been done. The pilot study cluster results were excluded from the final survey results; they were simply put aside and 25 'new' clusters selected. He had done this because the doctors and everyone else were inexperienced at the time; they had also not asked all the questions asked during the full survey. Only part of the required information had been collected during the pilot study.

Dr. Tamburi regretted that at the time of the pilot study he had not completed the design of his questionnaires and so was not able to test them during the pilot survey. This was something he would avoid during his next study; he felt it had been a bad mistake not to test his questionnaire before using it.

The last of these three difficulties is often the most troublesome. When it is too difficult to form small and more or less evenly sized clusters, then it is necessary to consider a two-stage sampling method.

Notes

TWO-STAGE SAMPLING

'Two-stage sampling' involves separate sampling in two stages. The basic method is easy and is done in four steps as follows:

The health problem leading to a survey

(i) Form clusters and select a random sample of these clusters (first stage sampling) as described under cluster sampling.

(ii) **For this sample of clusters only, make a separate list of all the study units within each of these clusters.** *No listing is needed for clusters not in the survey sample.*

(iii) Study units are now selected **separately** *from* **each** *of the clusters in the sample. These study units are chosen using list sampling or possibly by the 'numbered tag' method (second stage sampling).*

(iv) All selected study units are then examined or interviewed as required by the survey. Clusters and study units that have not been selected are not investigated.

Restricting the size of the survey

Two-stage sampling is a useful combination of list and cluster sampling. Its advantages are:

(i) It only requires the construction (making) of lists of study units for those

Typical sampling problems:

1. *How to select study units*
2. *How to obtain co-operation*

Sixth Speaker: Dr. Chan Yip

Subject: A Survey of Conditions in Restaurants and eating houses.

Dr. Chan referred to three severe outbreaks of food poisoning and gastro-enteritis which had resulted in about 80 hospital admissions and seven deaths. In all three outbreaks, the patients had reported eating at a restaurant or eating house. The City Health Centre was very concerned and doctors had for some time asked for food inspectors to be appointed to report on food preparation and general hygiene in the city's eating houses. After these outbreaks it was felt that they should again ask the City Authorities to appoint inspectors. Their request was more likely to be successful if they produced 'facts' and not just reported their impression that between 20% and 30% of eating houses were unhygienic.

To do a survey of the whole city of about one million people was far too difficult. But the patients in the three recent outbreaks had all reported eating in restaurants in a district called 'Riverside'. About one-third to one-half of all the city's eating houses were concentrated (located) in this part of the town.

The survey faced two main difficulties. The first was to get a sampling plan which would enable Dr. Chan to draw a random sample of eating houses in Riverside. There were not enough resources, and it also seemed unnecessary, to visit and inspect them all. The second difficulty was to get access (i.e. the right to enter and inspect) to the selected eating houses. The local police and health authorities would be able to issue a certificate to doctors giving them temporary authority to inspect eating houses. An inspecting doctor could also ask to be accompanied by a policeman during the inspection if he thought this necessary but only if he was refused access during his first visit. Every effort was made to avoid this arising. It was however decided that the inspecting

clusters drawn into the survey sample. For instance if only one-third of the clusters are selected, the work of listing will be reduced to about one-third of what it would be with list sampling. This is a great saving when lists are costly and difficult to make.

(ii) Only a sample of study units is chosen from each of the sample clusters. This means that the number of clusters in the sample can be increased without overburdening the survey with too many study units.

Practical reasons for choosing a two-stage rather than a simple cluster sampling plan

When using two-stage sampling, clusters can be made bigger and it is not as important to have all clusters of about equal size. This makes it easier to form clusters with easily recognisable and natural boundaries.

Equally sized clusters still have advantages, even in two-stage sampling, but the disadvantages of unevenly sized clusters are not so great.

In many applications of two-stage sampling the most difficult step is making a list of study units for those clusters drawn into the survey at the first stage of sampling. If lists of study units for these clusters exist, they should be used provided they are sufficiently up-to-date. However in practice it is unusual to find such lists and the survey organiser must arrange to make them.

Defining the survey area and the clusters within this area

Two separate sample size decisions

In a two-stage survey, the survey planner decides on two separate sample sizes:

(i) How many **clusters** *should be taken into the sample.*

Defining the survey unit

doctor should take along a male assistant from the Health Centre to give his visit more authority.

A two-stage sampling scheme was decided upon because:

(a) Cluster boundaries were unlikely to enclose approximately the same number of eating places.

(b) Because eating houses were very unevenly distributed, it was important to take a large sample of clusters to ensure a representative sample. This would however bring too many eating places into the survey if cluster sampling was done. Instead it was decided to take a sample of eating houses from *each* of the sample clusters, i.e. to use a two-stage sampling plan. This meant that for each of the sample clusters a separate list of eating places had to be made.

The definition of a study unit consisted of three parts:

(a) **Defining the survey area.** This was taken as the 'Riverside' area of the city as shown on the city street map.

(b) **Defining the clusters.** These were defined as the blocks (city areas) in Riverside bounded (enclosed) by three or more streets. On a copy of the city map, the Riverside area was divided into 317 blocks and numbered. Later it was seen that a few of these blocks contained no eating places, but one central block had as many as 22. Most city blocks had between 2 and 6 eating places.

(c) **Defining an 'eating place'.** For this survey the study unit was an eating place that satisfied the following three conditions:

1. Meals and cooked food are sold to the public and the eating place provides tables and chairs for eating this food on the premises. Eating places that sold cooked food and meals only for taking away without providing facilities (opportunities) for eating on the premises were excluded.

(ii) *How many* **study units** *should be taken from the sample of clusters. At least two study units should be taken but usually more than two are chosen from each sample cluster.*

Lack of Information

When planning a two-stage sampling scheme there is often not sufficient information to make the necessary sample size decisions.

More information is usually needed about:

1. *The size of the clusters. Some may be quite small, i.e. contain only a few study units, and some may be much larger. What appears to be a large cluster (in area) on the map may only contain a small number of study units.*

2. *In two-stage sampling it is usual to decide on a sampling fraction for selecting study units from within the sample of clusters. This sampling fraction determines the number of study units to be drawn into the survey. Unless the average or most usual cluster size is known it could happen that either too many or too few clusters are taken into the sample.*

For these reasons, as well as for questionnaire testing and interviewer training, a small pilot study is often undertaken to provide preliminary estimates (values) of the average cluster size.

The sample size. Decide on:

1. *the number of clusters needed*
2. *the number of study units needed*

2. Meals were sold and served to the general public on the premises at least five days in the week for most of the year. Eating places that had restricted membership (such as clubs and societies) were excluded.

3. Most of the food served had to be prepared and cooked on the premises or at least very close by, e.g. in the building next to the restaurant.

Two sample sizes had to be decided on:

1. How many clusters were to be taken.

2. How many study units, i.e. eating houses, to draw from each of the sample clusters.

Because of the uneven distribution (location) of eating places in Riverside, a large sample of clusters was needed to to be representative. A sample of 40 clusters was selected by drawing random numbers between 1 and 317. The remaining 277 clusters were not surveyed.

STUDY UNIT REPLACEMENT RULES

Most surveys encounter (meet) the problem of the 'not available' or 'will not co-operate' study unit. The need for a replacement rule arises whatever sampling plan is adopted, whether it is list sampling, stratified or any other kind of sampling scheme. The reason for discussing 'replacement rules' at this stage is that it was an important part of the survey of conditions at restaurants and eating houses.

'Not available' can arise in many ways: the following are a few examples:

Constructing lists of study units for each cluster in the sample

1. *Not living at the address given.*
2. *Not at home when interviewer came.*
3. *Refusing to answer a questionnaire or to attend a clinic for examination.*
4. *Changed use, e.g. a shop now empty or used as an office.*

Rule for dealing with doubtful study units

5. *During the visit or examination it becomes clear the study unit does not satisfy all the conditions for inclusion in the survey.*

What is the survey organiser to do in such cases? Guidelines have to be decided before the field workers start because they must know how to deal with such problems.

Typical sampling problems:

1. *Uneven size of study units*
2. *Replacement rule for rejected study units*

There are two basic alternatives:

Dr. Chan also felt that about 120 eating houses would be sufficient for showing the general conditions existing in Riverside. At this stage a final decision on the sample size was difficult because he did not know how many eating places existed in the sample clusters. This information would be available only after listing the eating places for each of the 40 sample clusters.

Listing the study units was done by a senior assistant from the Health Centre. This person walked around the blocks and into the buildings and made a list of all the eating places; he wrote down their name and address and any special features that would help the investigators find them. At this stage few questions were asked at the restaurant as otherwise improvements might be made and the place kept more clean than was normally the case.

The senior assistant was instructed to list all the eating places that *seemed* to meet the survey definitions. In doubtful cases he was still to list these eating places. The inspecting doctor would make the final decision whether an eating house was to be included.

Even at the early planning stages it was realised that the rules for drawing the second stage samples presented some problems. These problems were:

1. Eating places were very uneven in size; some restaurants were large whilst others were small, having only two or three tables. The selection had to ensure that the proper proportion of large and small eating houses all came into the sample.

2. Any list of eating places, however carefully done, would contain the names of eating houses which did not fully satisfy the survey definitions. Such eating places were to be omitted and a randomly selected replacement taken instead.

(i) Leave these study units out of the survey and **not** *replace them by other study units.*

(ii) Provide rules by which the 'not available' study units are substitued (i.e. replaced) by other **randomly** *chosen study units. It is important that the rules of replacement make it easy for the field worker to find a replacement study unit. These rules must also be clear and must allow the field worker no choice in selecting the replacement.*

The need for very specific sampling and selecting rules

After some discussion with others Dr. Chan set out the following sampling rules:

1. If a block had three or less eating places then all were to be included except those failing to meet the survey definitions.

2. If a block contained four or more eating places then the inspecting doctor was to inspect restaurants *in the order in which they were written* on the 'inspector's list'. The number of eating houses to be inspected was given at the top of this list.

3. If he found an eating place that did not satisfy the survey definition he was to take the *next* restaurant on his list. As an example, for block 163, the Five Star Club only served meals to its own members although it did allow special occasion dinner bookings by other societies and organisations. This did not meet the survey definition and the Five Star Club was excluded from the study.

There are many ways in which the field workers can be directed to the replacement study unit. A very useful method is to provide each field worker with a list of persons or places to be visited. This list has three essential parts:

1. *The name of the field worker and sometimes the date on which the investigation is to be done.*

2. *The number of persons or places the field worker is to visit or inspect.*

3. *A list of names and addresses of persons or places the field worker is to visit or inspect.* **There are always a few more names and addresses on the list than the field worker is required to visit.**

The field worker is instructed to go to each address on his list, **starting at the first one,** *until he has successfully interviewed or examined the number of study units he was instructed to visit. If any of the persons or places he visits is 'not available', he leaves it out and takes the* **next** *address down the list. The field worker finishes when he has successfully visited or interviewed the required number of study units,* **not one more and not one less.**

The organiser, in making 'interviewer lists', must do two things:

1. *All the names and addresses of study units on the list must be randomly drawn from the population by a proper sampling method such as described in this book. There must be a few more drawn than are required to be visited by the field workers.*

2. *The names and addresses of study units must be written on the list in random order. This can easily be done by writing out the list twice, first listing and numbering each address. Then draw random numbers and copy the names and addresses on to the second list*

Inspector's List

City Block 163

Enclosed by: Kwantoo and River Front Streets, China Road and Hibiscus Lane.

Number of Restaurants to be inspected: 4.

Inspecting Doctor: Dr. Tai Gan

1. Diner's Paradise	40 China Road
2. Sung Lung Restaurant	3 Kwantoo Street
3. The Five Star Club	13/14 Hibiscus Lane
4. Leetion's Late Night Eating Place	36 China Road
5. Lo Foo's Restaurant	51 China Road
6. Peking Eating House	57. River Front
7. The Riverside Hotel	21. China Road

in the same order as the numbers are drawn. **The second is given to the field worker;** *the first list is retained in the office and only used for checking purposes.*

When should a 'not available' study unit be replaced?

The following guidelines should be followed:

(a) **Replace by another study unit if:**

(i) the selected study unit does not meet the survey definitions, i.e. is not really a part of the study;

Typical problem:

Ensuring random replacement of rejected study units

(ii) the reason for not being available is completely unrelated to the survey and its aims. This is usually difficult to decide. **If unsure, then do not replace.**

Typical problem:

The final sample size is not the same as the sample size aimed for

(b) **Study units should not be replaced if:**

(i) The study unit is unco-operative or refuses to take part. Such study units are often different in attitude, social background and health from study units that do co-operate. Because of this the replacement study units are unlikely to be similar to the ones left out: it is best **not** *to replace them.*

Hence for block 163, eating places 1, 2, 4 and 5 were inspected but not numbers 3, 6 and 7.

Dr. Chan emphasised that the order of the restaurants on the inspector's list had been scrambled, i.e. randomly mixed up. This method allowed the inspector to leave out any eating houses that did not satisfy the survey definition but at the same time left him no choice as to which restaurant on his list was to replace it; it had to be the next one down.

The final survey sample contained 111 eating houses which was a little smaller than the 120 planned for. However 12 eating houses had been inspected or partly inspected before it became clear they did not meet the survey definitions. This meant that as far as survey resources of staff, time and money was concerned, a total of 123 eating houses had been investigated. This was as much as could be managed with the funds available.

The chairman (convener) of the session, Dr. Retlaw, after allowing a few questions, announced that the sessions had been very interesting and stimulating. He thanked all the speakers for their contributions. However the session had gone on much longer than intended. Because of this there was not time for his summing-up talk. The next day's session on 'Sample Size' was expected to be much shorter and he proposed to give the summing-up of this afternoon's discussions then.

(ii) **'Not at home' persons should not be replaced.** *Such persons should be revisited if resources and time allow. If after the second or third visit the interviewer still has not seen them, then they are best left out of the study and not replaced. A note should be made that they were not 'at home'.*

(iii) *'Cannot be found' or 'moved away' study units are likely to be different from those located (found). For this reason it is better not to replace them. For example, if a city list of shops is used for a survey, some of the selected shops may have closed down. These should not be replaced. A closed down shop is often an unsuccessful business, whereas a replacement shop is more likely to be successful. Frequent replacements of this kind will indicate better conditions than really exist.*

Whether a study unit is replaced or not, always make a note for later reference when a study unit presents some difficulty; note down what the difficulty is.

THE SAMPLE SIZE

The size of a survey, i.e. the number of study units visited, must be sufficient to make the survey results reliable. In general, a large study contains more information and will provide more reliable results (answers).

The size of a survey is usually a compromise between the need for reliable results, which require a large sample, and keeping the study small enough to be done by the available staff and resources.

Deciding the Sample Size

Deciding the sample size occurs in two stages:

1. *First decide on the single most important proportion the study is to measure. Then find a preliminary estimate (informed guess) as to what this value is likely to be.*

2. *Next, use this preliminary estimate of the most important proportion to find the required sample size with the help of the two tables given below. This second stage has to be applied in different ways depending on the sampling method used.*

Session 3: Survey Workshop

Convenor: Dr. L. Retlaw

Subject: The Sample Size

To start the session, Dr. Retlaw put two statistical tables up on display, Table 1 and Table 2, which would be needed to determine the sample size. The speakers would refer to these when explaining how they had calculated their survey sample size.

Dr. Retlaw emphasised that the sample size as calculated from these tables would not always be accepted because other practical considerations made this impossible. However it was most important to calculate the desirable sample size as this indicated the sample size at which to aim.

Finally, Dr. Retlaw wrote the definition of a study unit on the blackboard; he felt it was important for everyone to understand its meaning.

'**Study Unit**: this is the basic or smallest 'unit' with which the survey is concerned. It is this unit that the field workers must ultimately visit for interviewing, for inspection or for study.'

SAMPLE SIZE FOR LIST SAMPLING

For list sampling, use the preliminary estimate of the proportion and enter Table 1 to read off the corresponding desirable sample size. If the preliminary estimate is given as a minimum and maximum value within which the final survey value is likely to be, then look up the sample sizes required for both the minimum and the maximum values. The **larger** *of these two sample sizes is used.*

Two strata formed

Preliminary estimate of proportion

TABLE 1

Sample Size for List Sampling

Estimated Proportion	Desirable Sample Size*	Estimated Proportion
0.05	420	0.95
0.10	325	0.90
0.15	290	0.85
0.20	255	0.80
0.25	225	0.75
0.30	195	0.70
0.35	170	0.65
0.40	145	0.60
0.45	120	0.55
0.50	100	0.50

How to use Table 1

The desirable sample size is given in the middle (second) column. The table is entered using either the left (first) column or the right (third) column depending on whether the estimated proportion is less than or greater than 0.5.

* For the information of survey specialists:
In this table the estimated S.E./p gradually increases from 0.10 for p = 0.5 to 0.21 for p = 0.05.

First Speaker: Dr. K. Singh

Subject: 'A Survey of School Performance: The Sample Size'

Dr. Singh emphasised that his survey plan for measuring school performance depended upon forming two strata. One strata consisted of a list of 816 boys who satisfied the survey definition. The second strata was a list of 471 girls.

The principal aim of this survey was to measure the proportion of children receiving help at home. There was no information on this. Dr. Singh assumed that most literate (able to read and write) parents would help their children with their schoolwork. About 40% of families had at least one literate parent. It was therefore assumed that about 40% of children would receive help and this was taken to be the preliminary estimate, i.e. a proportion of 0.4 received assistance at home.

From Table 1, it is seen that the desirable sample size for a proportion of 0.4 is 145; this was rounded up to 150. This then was the size of sample to be taken from *each* of the two strata.

The survey was intended also to test whether fewer girls received help at home or received less help than did boys. Some teachers felt that only about 20% (0.2) of girls were assisted with school work. If this were true then a sample size of 255, as can be seen from Table 1, should be taken from the list of girls. However, as explained during the morning session, there were not enough resources to interview and test more than 150 girls.

Notes

Sample Size and Sampling Fraction

To determine the sampling fraction for a 'numbered tag' survey plan, it is necessary to proceed as follows:

Step 1 *Obtain a preliminary estimate of the most important proportion.*

Or

Obtain an estimate of the lower and upper limits within which this proportion is likely to lie.

Step 2 *Use Table 1 to find the required sample size (as for list sampling).*

Step 3 *Estimate the number of study units going through the clinic (i.e. the number of study units involved).*

Step 1 Applied

Estimating likely limits of the proportion

Step 2 Applied

Finding the desirable sample size

Step 3 Applied

Estimating the number of patients involved

Calculating the sampling fraction

Finally, the required sampling fraction is obtained by dividing the required sample size by the number of study units involved.

Second Speaker: Dr. A. Xuma

Subject: 'Calculating the Sampling Fraction for a Survey of Parasitic Infestation at a Child Care Clinic'

Dr. Xuma reminded the audience that the level of infestation was put as low as 10% by himself or as high as 75% by another physician, i.e. a proportion likely to be between 0.1 and 0.75. From Table 1 the corresponding sample sizes were 325 and 225. Of these, the higher value of 325 was accepted as the desirable sample size.

The study was planned to continue over 30 sessions. Between 50 to 70 infants between 6 months and 3 years were likely to be seen each session; an average of 60 per session. Over the 30 session period this comes to about 1800 children. The sampling fraction is then $325/1800 = 0.18$, or between 1 and 5 or 1 in 6.

For various practical reasons such as waiting room space and laboratory resources, the sampling fraction used in the survey was 1 in 10, which gave an actual survey sample size of 189 infants. This was much less than the desired size of 325.

Despite this smaller survey, the main conclusion of the survey was reliably established. The young doctor was clearly mistaken in thinking the infection level was close to 75% but at the same time the survey strongly indicated that the level was about 20% and this was disturbingly high.

SAMPLE SIZE AND SAMPLING FRACTION FOR STRATIFIED SAMPLING:

For surveys using stratified sampling it is necessary to carry out similar but separate studies in each strata. The same questionnaires are used for all the strata, the interviewers are the same persons and they are all given similar training; all the field work is done for all the strata during the same survey period.

The sample size needed for **each** *strata found as for list sampling, in two stages.*

(i) For each strata find a preliminary estimate, i.e. minimum and maximum likely value, for the most important proportion with which the survey is concerned.

(ii) Use Table 1 to find the sample size for each strata corresponding to these minimum and maximum proportions. If resources allow, for each strata choose the larger of these sample sizes.

Preliminary estimates

Sample size derived from Table 1

A study which aims to compare proportions from different strata requires to know these values reliably. Sample size requirements for comparisons can become very large. Because of this, survey planners sometimes accept the smaller of the sample sizes as given by Table 1. The results of the survey will still be acceptable but the results will not be as reliable as they would be using the larger sample sizes.

Third Speaker: Dr. J. Desai

Subject: 'The Sample Size required for an Accident Treatment Survey using Stratified Sampling'

Dr. Desai referred to the three Health Centre surveys he had described during the previous session. These three Centres served different geographical areas and communities. An important aim of the survey was to determine (measure):

(i) the proportion of accidents/injuries amongst persons presenting at the clinics;

(ii) the proportion of accidents/injuries occurring at work including fishing and agricultural activities.

Several comparisons were required such as whether the kind of accident/injury and their treatment differed between the three Health Centres.

Reliable preliminary estimates of the proportions were not available but it was thought that between 10% and 20% of presenting cases were due to accidents and injuries. About 40% of these accidents/injuries were thought to have occurred at work.

From Table 1 it is seen that for a proportion of about 0.4 (i.e. 40%) a sample of 145 cases is needed.

The 'Numbered Tag' sampling scheme used at the Health Centres produced large samples of a cross-section of presenting patients as shown below. On available infomation, at least 10% of these would be due to accidents and injuries. As the figures show, at all three Centres the number of accident cases selected and reviewed by this sampling scheme in a year exceeded the number required, i.e. exceeded 145 cases.

If numbered tag sampling is suitable then the sampling fraction is estimated for each strata separately. Each of these sampling fractions is determined as described under 'numbered tag sampling'.

Sampling scheme provides adequate sample size

		Total selected by Numbered Tag sampling	Minimum (10%) number of accidents expected to be included
Strata:	Centre 1	4113	411
	Centre 2	3602	360
	Centre 3	2130	213

Notes

SAMPLE SIZE FOR CLUSTER SAMPLING

There are several ways of deciding the sample size required for cluster sampling. The method given here is easy to apply and will give reliable results under ordinary survey conditions.

Finding the desirable sample size for cluster sampling is done in three simple steps:

(i) count the number of clusters formed, i.e. the total number of clusters into which the survey population has been divided;

(ii) make a preliminary estimate of the likely value of the most important proportion with which the survey is concerned;

*(iii) using this estimate of the most important proportion, look up the desired cluster sample size in Table 2.**

If the required proportion is stated as lying between some upper and lower limit, then the above procedure is used twice, once using the likely upper value and once using the likely lower value. The larger of the cluster sample sizes is accepted.

* *Instructions how to use Table 2 are given on page 100.*

Fourth Speaker: Dr. T. Tamburi

Subject: 'Finding the Sample Size for a Survey using Cluster Sampling'

Dr. Tamburi reminded the audience that his survey aimed at establishing the living conditions, the levels of T.B. and of eye infections in a town of about 40,000 people. During the morning session he had described how the town had been divided into 432 clusters, a sample of which were to be completely covered (every household visited) by a doctor and a Social Science Department interviewer.

It was important to measure the proportion of households in which there was at least one patient suffering from T.B. or infectious eye disease. The pilot study had given a preliminary proportion of 0.18.

For this survey, the first row (line) of Table 2 (page 96) applies because the total number of clusters formed is 'over 400'. The preliminary proportion lies between the '0.15' and the '0.20' columns; the '0.15' column is chosen because it indicates the larger sample sizes. The sample size at the intersection (crossing) of this column (the '0.15' column) and the first row is 23.

To be cautious, Dr. Tamburi decided on a slightly larger sample of 25 clusters.

Dr. Tamburi again mentioned his concern that the clusters formed should be small. From Table 2, it can be seen that if the survey population is divided into fewer large clusters, then the number of clusters needed in the sample decreases. However, the bigger the clusters, i.e. the more study units the clusters contain, the greater the survey field work becomes.

Dr. Tamburi went on to demonstrate how important

Table 2
Sample Size for Cluster Sampling

Total Number of Clusters Formed	Proportion Expected									
	.0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
over 400	26	24	23	21	20	18	17	15	14	13
400	25	24	22	21	19	18	17	15	14	13
300	25	23	22	21	19	18	16	15	14	12
250	24	23	22	20	19	17	16	15	14	12
200	24	23	21	20	19	17	16	15	13	12
150	23	22	21	19	18	17	16	14	13	12
100	22	20	19	18	17	16	15	14	13	12
90	21	20	19	18	17	16	15	14	12	11
80	20	19	18	17	16	15	14	13	12	11
70	20	19	18	17	16	15	14	13	12	11
60	19	18	17	16	15	14	14	13	12	11
50	18	17	16	15	15	14	13	12	11	10
40	16	16	15	14	14	13	12	12	11	10
35	15	15	14	14	13	12	12	11	10	10
30	14	14	13	13	12	12	11	11	10	9
25	13	13	12	12	11	11	10	10	9	9
20	12	12	11	11	10	10	10	9	9	8
15	10	10	9	9	9	9	8	8	8	7
10	8	8	7	7	7	7	7	7	6	6

Continued on page 98.

this was for figures obtained from another study he had done some years ago. This was a survey of a community close to 3,000 families. In that community about one family in five had an elderly parent or parents living with them, i.e. a proportion of 0.2.

In that community if small clusters are formed of about 10 families in a cluster, then the final desirable size of the study as obtained from Table 2 is:

Families	Approximate Number of Clusters formed	Cluster Sample Size from Table 2	Approximate Number of families in the survey
3,000	300	21	10 x 21 = 210

On the other hand, if that community is divided into 50 large clusters of about 60 families each, the survey has to deal with a smaller sample of clusters but with more families, as is shown below:

Families	Approximate Number of Clusters formed	Cluster Sample Size from Table 2	Approximate Number of families in the survey
3,000	50	15	15 x 60 = 900

This clearly shows the advantage of forming small clusters. A community divided into only a few large clusters requires a large survey for reliable results, large in terms of the study units coming into the survey.

Table 2 (continued)
Sample Size for Cluster Sampling

Total Number of Clusters Formed	Proportion Expected									
	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
Over 400	13	14	15	17	18	20	21	23	24	26
400	13	14	15	17	18	19	21	22	24	25
300	12	14	15	16	18	19	21	22	23	25
250	12	14	15	16	17	19	20	22	23	24
200	12	13	15	16	17	19	20	21	23	24
150	12	13	14	16	17	18	19	21	22	23
100	12	13	14	15	16	17	18	19	20	22
90	11	12	14	15	16	17	18	19	20	21
80	11	12	13	14	15	16	17	18	19	20
70	11	12	13	14	15	16	17	18	19	20
60	11	12	13	14	14	15	16	17	18	19
50	10	11	12	13	14	15	15	16	17	18
40	10	11	12	12	13	14	14	15	16	16
35	10	10	11	12	12	13	14	14	15	15
30	9	10	11	11	12	12	13	13	14	14
25	9	9	10	10	11	11	12	12	13	13
20	8	9	9	10	10	10	11	11	12	12
15	7	8	8	8	9	9	9	9	10	10
10	6	6	7	7	7	7	7	7	8	8

How to use Table 2

1. *First count the total number of clusters formed. Use this total to find the corresponding row (line) of Table 2 (see first column). If this total does not appear in the first column, take the nearest bigger value. For example, if there is a total of 85 clusters, use the row with '90' as being the nearest bigger value in Table 2.*

2. *Next, see which of the proportions in the column headings of Table 2 is closest to the estimated proportion. If the estimated proportion falls between two columns, use the column with the larger sample sizes.*

3. *Finally, the required cluster sample size is given at the intersection (crossing) of the row and column found under (1) and (2) above.*

Note:
Occasionally some sample clusters are found to contain no study units. **Such 'empty' clusters must not be replaced but must remain part of the survey results.**

Notes

SAMPLE SIZE FOR TWO-STAGE SAMPLING

For two-stage sampling, two separate sampling decisions are needed. These are:

(i) How many clusters should be taken into the survey sample.

(ii) How many study units should be taken from the chosen clusters. This is best done by using the same (fixed) sampling fraction for each cluster in the sample.

The following guidelines should be borne in mind:

Typical Problem:

Resources not sufficient for the desirable sample size

(i) **For cluster sampling it is more important to increase the number of clusters chosen than it is to increase the number of study units chosen from the clusters.** *For example, a sample of 40 clusters with four to five study units drawn from each provides the survey with between 160 to 200 units. A survey of about the same size will result if only 20 clusters are chosen but twice as many study units are taken from each. However the first sampling scheme, i.e. taking a sample of 40 clusters, is the better plan.*

(ii) **If possible, arrange for at least three study units to be drawn from each cluster in the sample.**

The sample sizes needed for two stage sampling are best determined in four simple steps:

1. *Estimate the value of the most important proportion, or its likely upper and lower limits. Use Table 1 to decide on the* **total** *sample size (for study units). If this*

Fifth Speaker: Dr. Chan Yip

Subject: 'Sample Sizes for a two-stage Sampling Survey of Restaurants'

Dr. Chan Yip said he would be very brief and just quickly go through the steps taken to decide on the sample sizes in his survey of restaurants in Riverside.

From previous experience the proportion of unhygienic or otherwise unsatisfactory eating places was thought to lie between 0.2 and 0.3.

The corresponding sample sizes taken from Table 1 were 255 and 195. This was beyond the resources of the study and it was decided to restrict the sample to 120 eating places.

The next step was to look up Table 2 to obtain the least number of sample clusters required; Riverside has been divided into 317 clusters (blocks).

The nearest (higher) total in Table 2 is 400 and this was the row used. For the columns of Table 2, the proportion of 0.2 was used, rather than 0.3, because this required the larger sample. Using these values, Table 2 showed the minimum cluster sample size to be 21.

Dr. Chan felt uneasy about such a small sample of city blocks because restaurants are very unevenly distributed in Riverside. Also eating places differ greatly in size and character. For this reason he increased the sample size to 40 clusters.

The preliminary listing of restaurants in these 40 blocks revealed 257 eating places. The number of eating places to be surveyed was 120.

The sampling fraction was then calculated as:

sample size is too large for the available survey resources, **decide upon the largest number of study units the resources allow.**

2. **Use Table 2 to decide the least number of clusters needed.** *This requires:*

   (a) the estimated proportion or its likely upper and lower limits as determined in (1) above;

   (b) the number of clusters formed.

   For two-stage sampling it is good practice to increase this cluster sample size by at least 20%.

   Empty sample clusters included

3. *Determine the number of study units in the sample of clusters by listing them.*

4. *The required sampling fraction is then given by dividing the number of study units required by the survey (as found in (1) above) by the total number of study units in the sample of clusters (as found in (3) above).*

Occasionally clusters will be 'empty', i.e. contain no study units. **Such empty clusters, once they have been drawn into the sample must be retained (kept) and must not be replaced by choosing another cluster.** *The sample size still remains equal to the number of clusters selected initially, including the number of empty clusters.*

$$\frac{120}{257} = 0.467 \qquad \text{or} \qquad \text{about 1 in 2.}$$

This was the reason for taking a random selection of half the study units from each sample cluster, rounding up if 'one-half' was not a whole number.

Where possible at least three eating places per cluster were selected. All eating places were visited in those sample clusters that contained three, or less, restaurants.

Two of the sample of 40 clusters were found to have no eating places. These were **not** replaced by other clusters. In all his later calculations of the survey results Dr. Chan always worked with a sample size of 40 clusters and not with 38, i.e. the two empty clusters continued to be included in the analysis.

SURVEY SAMPLING

SUMMARY AND REVIEW

Most small scale or local health surveys can be done using one of five basic sampling schemes. These are:

1. *List sampling*
2. *Numbered Tag sampling*
3. *Stratified sampling*
4. *Cluster sampling*
5. *Two-stage sampling*

Review of Terminology and Definitions

1. **Study Unit**

This is the basic or smallest 'unit' with which the survey is concerned; it is this unit that the field workers must ultimately visit for interviewing, inspection or study.

2. **List Sampling**

This consists of making a numbered list of every study unit in the survey population and then drawing a random sample of these study units for inclusion in the survey.

3. **Numbered Tag Sampling**

This consists of issuing (giving) every person a numbered tag as he comes to a clinic or applies for some service. Only persons whose tag number ends in previously agreed digits (numbers) are included in the study.

4. **Stratified Sampling**

The whole survey population is divided into groups or 'strata' in such a way that within each strata the study units are more alike than they are in the survey

Sixth Speaker: Dr. L. Retlaw

Subject: 'Review of the Day's Discussions'

Dr. Retlaw started by congratulating the speakers on their papers and the interesting survey problems they had discussed. There were two aspects on which he would like to expand. The first of these was the problem of obtaining a preliminary estimate of the most important proportion the survey aimed to measure.

Often it is enough to decide that this proportion lies between a lower (minimum) and an upper (maximum) value. The three ways of getting the preliminary estimates of these values are:

1. Using results from other, but similar studies. These may have been done elsewhere or done on an earlier occasion.

2. Doing a pilot study to obtain these preliminary estimates.

3. Using local experience and knowledge as to what might be expected. Where several preliminary proportions are suggested, the sample size is estimated for **each** of these. The largest of the sample size estimates is then used.

The closer the preliminary estimates are to the final results of the survey, the better. However, little harm is done if the preliminary proportions are not very close to the final survey results.

The second aspect Dr. Retlaw emphasised was that of relating the desirable sample size requirements to available survey resources.

Sample size rules are to be taken as guidelines; they do not have to be followed completely. If resources are in-

population as a whole. Separate areas or institutions in which similar social or health conditions exist, can also be considered as 'strata'. These strata must cover the whole population. A separate sample is taken from **each and every** *strata using the list sampling procedure.*

5. **Cluster Sampling**

This consists of groups or clusters of sampling units enclosed in an easily recognisable boundary. In forming clusters, the study units within a 'cluster' do not need be similar. As far as is practical, all clusters should contain approximately the same number of study units. A random sample of these clusters is chosen by list sampling and **all** *the study units within the selected clusters are examined or interviewed.*

6. **Two-stage Sampling**

Clusters are formed as for cluster sampling and a random sample of these clusters is chosen. A list is then made of all the study units within the selected (sample) clusters. By the list sampling method, a sample of study units is then drawn from each of the selected (sample) clusters.

Advantages and Limitations of the Different Sampling Methods

1. List Sampling

This sampling method has great simplicity and is recommended whenever lists for the whole survey population are available or can be constructed easily. Available lists are sometimes incomplete and out of date. If lists are unavailable or costly then either cluster sampling or two stage sampling should be considered.

sufficient, then a smaller survey can be done. However, the survey planner must realise and accept that the smaller his survey, the less reliable are the survey results (answers) in general. It is then necessary to decide whether the results will be sufficiently reliable to make a survey worthwhile. Alternatively, the planner can try to reduce the number of questions asked during interviewing or reduce the number of investigations and tests done. This reduces the scope (extent) of his study and so lessens the cost per study unit examined. This saving will allow a larger sample to be taken.

Dr. Retlaw went on to say that only exceptionally should a conflict between available resources and sample size deter (prevent) the survey planner from doing any study at all. Even when the resources allow only a small study, much smaller than indicated by the desirable sample size, there are still three important advantages in doing a survey:

1. Planning a survey, however small, drafting questionnaires, talking to patients and to families in the area always reveals new aspects and teaches new insights into problems. The survey organiser is always better informed and more experienced **after** a survey than he was **before** he started.

2. The experience gained by nurses and field workers from talking to people, investigating ill-health and other problems results in a better appreciation of community needs. This experience will also be helpful to subsequent (later) studies.

3. The results of a small survey, even if not entirely reliable, can be used as a pilot study. These results help to plan a bigger survey and provide the important information required when seeking financial support for a larger investigation.

2. Numbered Tag Sampling

This type of sampling should be considered whenever there is a flow (steady arrival) of people (e.g. patients or shoppers) who arrive or report to obtain help, treatment, advice or some other service. At the place of service or at the time of treatment, it must be possible to determine (test) whether such a person belongs to the survey population or not, i.e. satisfies the survey definition of a study unit. If this cannot be done then numbered tag sampling is not practical.

3. Stratified Sampling

Stratified sampling should be considered whenever:

(i) it is possible to divide the survey population into two or more strata comprising (containing) somewhat similar study units or enclosing areas in which similar conditions exist;

(ii) it is not too difficult to construct a list of the sampling units in **each** *of the strata.* **Stratified sampling permits easy comparisons between the strata.**

4. Cluster Sampling

Cluster sampling almost makes listing unnecessary; only a list of the clusters is needed. For this reason cluster sampling is attractive wherever the listing of the study units is difficult. To ensure a representative sample an adequate number of clusters must be selected. However this may result in too many study units coming into the survey unless the clusters are small and of similar size.

A Survey Planner should not be too easily discouraged.

Finally, Dr. Retlaw stressed the need for planners and organisers to become familiar with the survey area. **They needed to visit it and to see conditions for themselves.** It is not sufficient to send an assistant to the area or to the institutions to be surveyed. Planners and organisers must get to know the difficulties their interviewers will face. They must take account of these difficulties in their planning and later in the design of the questionnaires and in the training of the field staff. He again stressed the importance of planners visiting the area to see conditions for themselves **before** planning started. A proper 'site' (area) visit could take from a day to a week or more.

In conclusion, Dr. Retlaw referred to information boards he had placed outside the lecture room. These boards showed flow charts that indicated, in simple steps, how to choose the most appropriate sampling scheme.

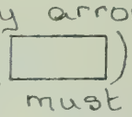
5. Two-stage Sampling

Two-stage sampling should be considered whenever it is costly or difficult to list the whole survey population. Two-stage sampling reduces the amount of listing required whilst at the same time overcoming the disadvantages of cluster sampling. In particular it allows the survey planner to take a larger number of clusters into his sample without overburdening the study with too many study units.

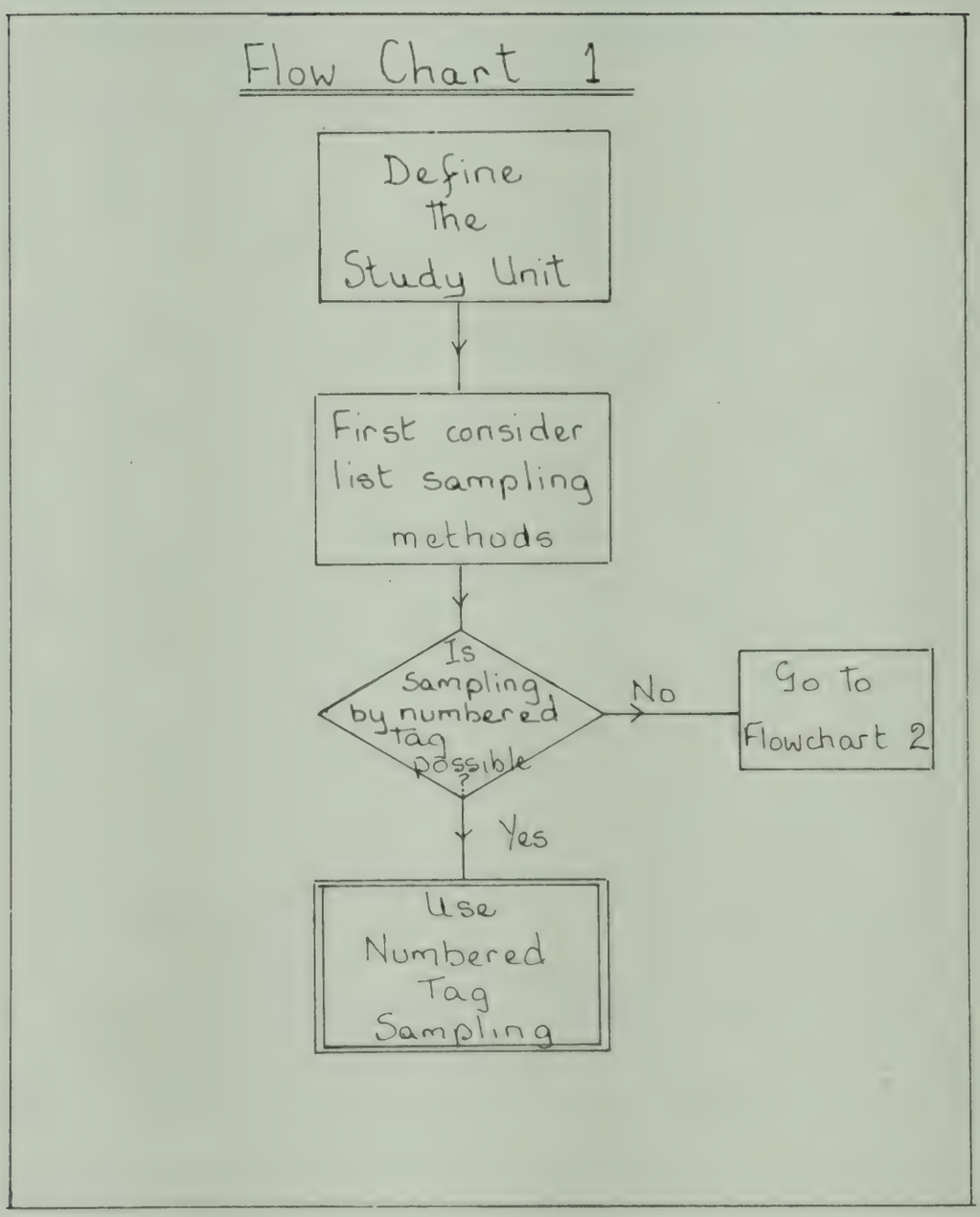
The Exhibition of Information Boards

1st Board

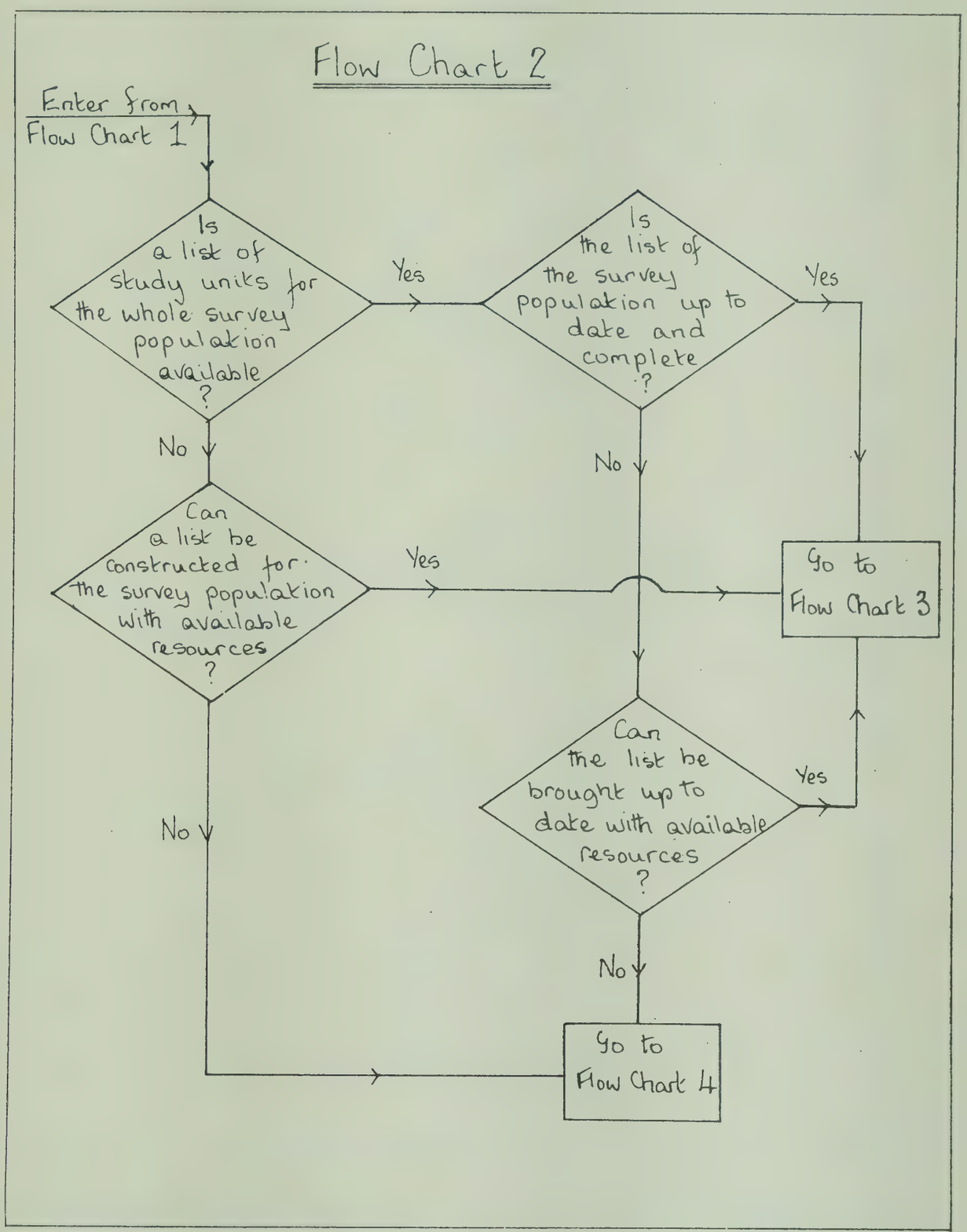
Using the Flow Charts

The flow charts consist of a sequence of boxes joined by arrows (pointers). Rectangular shaped boxes () contain instructions the survey planner must carry out before going further along the flow chart. Diamond shaped boxes () contain questions which the user answers with a simple 'yes' or 'no'. If the answer is 'yes', the planner follows the pointer marked 'yes' to the next box, otherwise he follows the line marked 'no' to another box. This procedure always ends up with advice to the user as to which of the sampling methods is best suited to his survey problem. The sampling scheme which is recommended is shown in a double lined rectangular box.

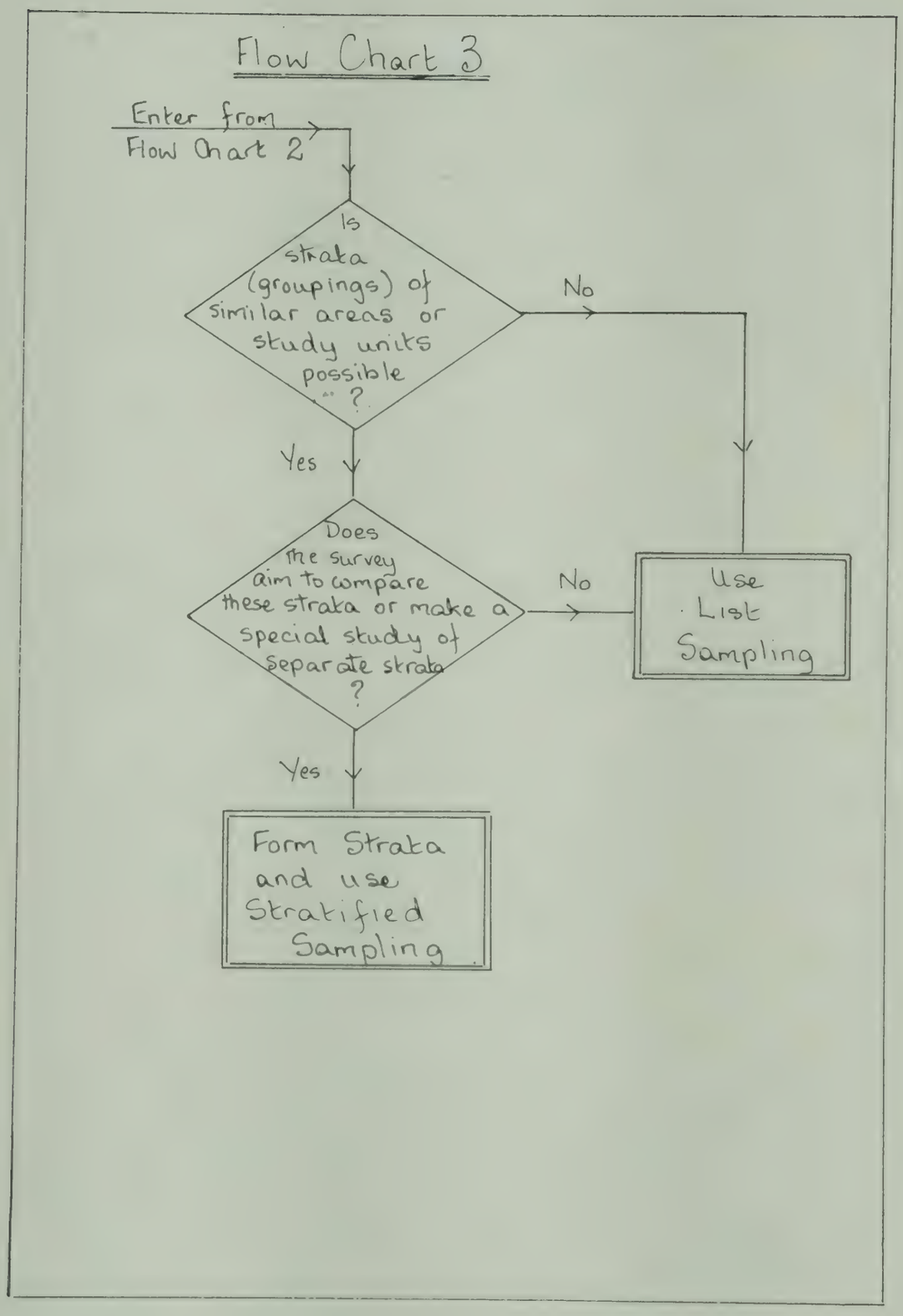
2nd Board


Flow Chart 1
Define the Study Unit
First consider list sampling methods
Is sampling by numbered tag possible?
No
Go To Flowchart 2
Yes
Use Numbered Tag Sampling

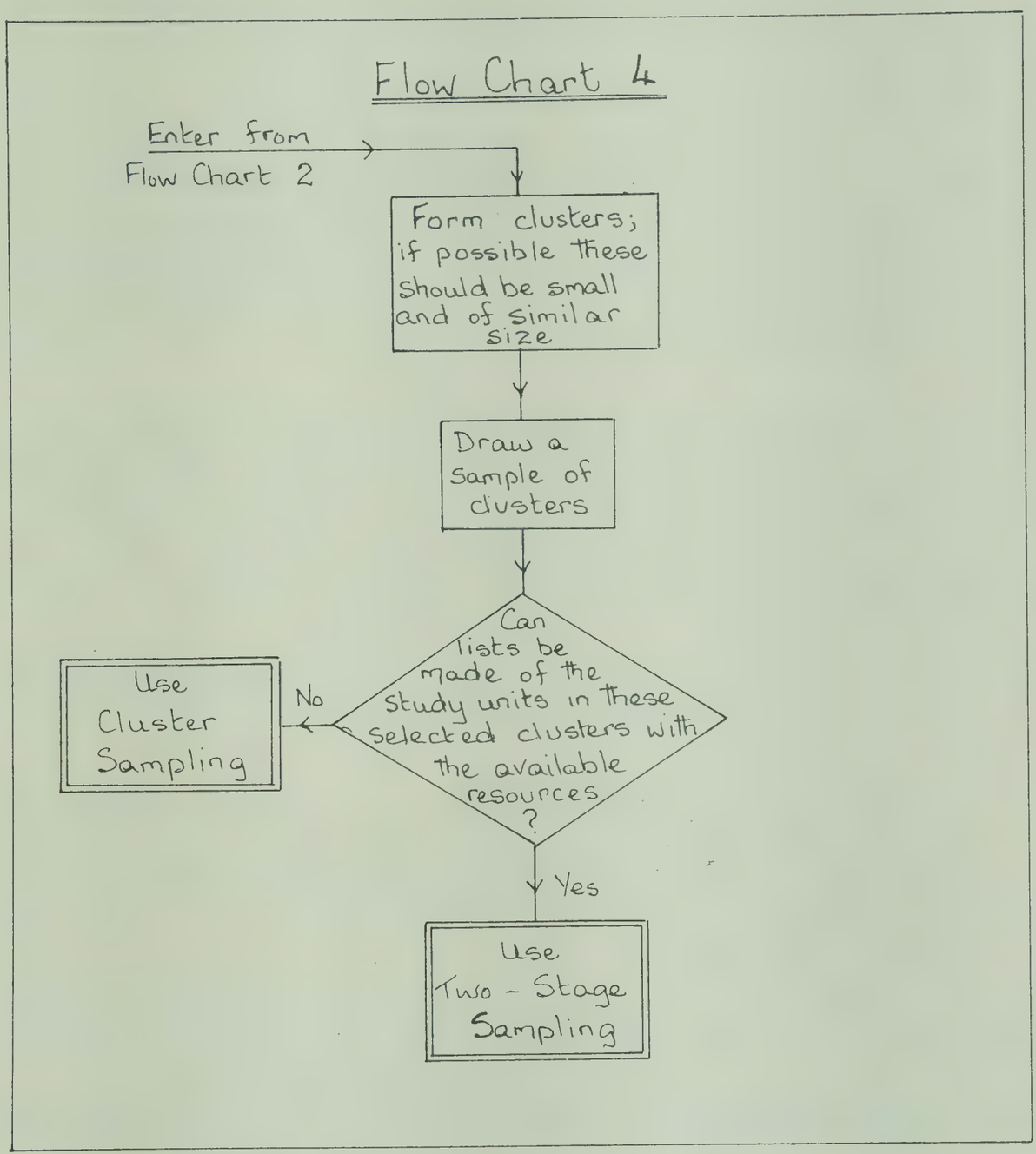
3rd Board


Flow Chart 2
Enter from Flow Chart 1
Is a list of study units for the whole survey population available ?
Yes
Is the list of the survey population up to date and complete ?
Yes
No
No
Can a list be constructed for the survey population with available resources ?
Yes
Go to Flow Chart 3
Can the list be brought up to date with available resources ?
Yes
No
No
Go to Flow Chart 4

4th Board


Flow Chart 3
Enter from Flow Chart 2
Is strata (groupings) of similar areas or study units possible ?
No
Yes
Does the survey aim to compare these strata or make a special study of separate strata ?
No
Use List Sampling
Yes
Form Strata and use Stratified Sampling

5th Board


Flow Chart 4
Enter from Flow Chart 2
Form clusters; if possible these should be small and of similar size
Draw a sample of clusters
Can lists be made of the study units in these selected clusters with the available resources ?
No
Use Cluster Sampling
Yes
Use Two - Stage Sampling

Appendix

DRAWING A SAMPLE OF RANDOM NUMBERS

1st Method: Selecting numbers from a box

Cut sheets of paper into squares large enough for writing numbers, usually four or five centimeters square (2 inches). On the first piece write '1', on the second piece '2', on the third '3' and so on up to the last number on the survey list. Then fold each paper separately and tightly so that the number cannot be seen. Put all these folded papers into a box and **mix (stir) them up very well several times.** *When this has been done, let someone else draw out as many of these folded papers as the sample requires.*

Unfold the drawn (sample of) papers and select from the survey list those names whose number is the same as that on the selected paper.

Points to Remember

(a) Each name on the list has one number and this number is not used against any other name.

(b) Each folded paper has only one number on it; the numbers on the paper must correspond to the numbers on the list. There must be no additional numbers and no numbers left out.

2nd Method: Selecting from a Table of Random Numbers

How to use a Random Number Table

Most random number tables consist of blocks of five numbers of five digits each (see Table 3, page 123). These numbers may be read in any order, up or down the columns or along the rows. Numbers are selected from this table and

the name on the survey list with the **same** *numbers is chosen for inclusion in the study. Sometimes the same random number comes up a second or more times but it is then ignored. Sometimes the number read from Table 3 is greater than the largest number on the list; this number is also ignored.*

How to use Table 3

Step 1

Determine the number of digits in the largest number on the survey list. For example, if 317 blocks (clusters) are formed, the largest number is 317 and it has three digits.

Step 2

Table 3 contains five digit numbers which are bigger than needed for most surveys. These numbers can be reduced to smaller numbers by ignoring unwanted digits. For example, if a three digit number is wanted then the first number in Table 3, which is 44983, can be read as a three digit number in several ways as follows:

(i) as 449 (ignore last two digits)

Or

(ii) as 498 (ignore first and last digit)

Or

(iii) as 983 (ignore first two digits).

All three numbers are equally valid (usable) as three digit random numbers.

Step 3

Having chosen a number with the right number of digits from Table 3, select the corresponding item on the survey list for inclusion in the study, unless:

(i) the number chosen has been drawn (selected) previously;

(ii) the number chosen is greater than the largest number on the list.

Step 4

Take the next number in Table 3 and proceed as in Steps 2 and 3. The next number can be chosen **anywhere** *in Table 3. It is easiest to go down the columns, one after another until the number of study units chosen from the survey list equals the sample size required. As a precaution, some statisticians select the starting row and column randomly; this practice is recommended.*

Note

Random numbers such as 007 and 0173 are read as 7 and 173, i.e. **zeros at the start of a number are ignored.** *Zeros within a number, such as 103, or at the right end, such as 270, are of course* **not** *ignored, and continue to have their ordinary meaning.*

TABLE 3 Random Numbers

44983	33834	54280	67850	96025	96117	00768	14821	69029	25453	48798	15486
89494	34431	44890	59892	79682	20308	82510	53609	13258	89631	80497	49167
54430	52632	94126	95597	48338	67645	44676	14730	22642	21919	21050	87791
96999	42104	34377	63309	82181	00278	28209	95629	75818	09043	48564	87355
87947	09427	32380	43636	58578	07761	28456	46570	11623	50417	37763	30136
30238	46126	85306	37114	22718	50584	92291	56575	24075	43889	40909	18741
22938	13073	32066	43098	75738	94910	15403	89151	73322	18370	90586	46115
89182	27750	63314	87302	49472	24885	79506	60638	07132	00908	92035	75518
16187	03303	40287	52435	23926	92544	54099	31497	06863	22864	72620	74169
21526	07401	30925	46148	20138	33874	56715	38424	38273	11361	15203	64912
42907	95158	27146	37012	43361	03173	97911	71313	44256	66609	42504	76799
21479	48265	01674	47274	56350	37512	14883	99673	62298	33948	32456	28675
90076	70233	76730	25043	16686	54737	57431	01786	20803	69465	37970	05673
93202	25355	93941	84434	22384	13240	93617	51549	28532	57150	77261	62643
46059	72208	90475	10341	39703	83224	37858	61657	04184	15597	29448	01922
38220	13972	86115	17196	24569	26820	66299	39960	02489	53078	72789	22562
82618	85756	51156	74037	12501	94162	42006	16135	82797	31296	93268	10104
07896	74085	59886	03051	78702	13402	74318	10870	72107	11550	61175	33345
95241	84360	13960	95736	43637	60399	19080	60261	11207	73065	48286	57057
53849	26578	39954	86726	91039	13884	25376	36880	02564	96978	62332	77321
72967	53031	47906	99501	27753	69946	66875	25601	30038	78786	65197	65283
87910	89260	66444	15979	83469	76952	50065	72802	70630	87336	16385	32784
10482	34277	40177	01081	57788	08612	39846	42234	04905	83274	22459	75032
68034	98561	46747	30655	41878	93610	51745	41771	61398	98154	61644	12405
80277	92450	60888	18689	45966	25837	70906	60733	11765	09293	70076	40751
59896	78185	60268	03650	36814	88460	34049	09111	64205	77930	32391	69076
78369	04163	77673	73342	78915	20537	06126	27222	17378	59359	00055	66780
23015	54261	95020	77705	81682	96907	37411	93548	87546	07687	47338	12240
55171	85448	12545	75992	08790	88992	69756	18960	85182	02245	11566	52527
58095	62204	69319	00672	96037	78680	98734	83719	40702	79038	68639	63323
19700	98193	37600	70617	58959	45486	58338	84563	62071	17799	96994	41635
12666	87597	23190	26243	36690	75829	71060	32257	15699	02654	83110	44278
66685	05344	71633	68536	18786	28575	08455	79261	49705	31491	25318	52586
72590	47283	45445	35611	98354	53680	45747	62026	13032	14048	16304	11959
30286	06434	50229	09070	44848	09996	77753	05018	92605	10316	07351	78020
87494	95585	25547	53500	45047	08406	66984	63390	48093	02366	05407	08325
32301	25923	76556	13274	39776	97027	56919	17792	09214	53781	90102	25774
70711	37921	54989	17828	60976	57662	61757	93272	09887	34196	98251	52453
36086	05468	41631	95632	78154	38634	47463	37514	24437	01316	04770	06534
37403	42231	17073	49097	54147	03656	14735	06370	18703	90858	55130	40869
41022	76893	29200	82747	97297	74420	18783	93471	89055	56413	77817	10655
70978	57385	70532	46978	87390	53319	90155	03154	20301	47831	86786	11284
19207	41684	20288	19783	82215	35810	39852	43795	21530	96315	55657	76473
50172	23114	28745	12249	35844	63265	26451	06986	08707	99251	06260	74779
43112	94833	72864	58785	53473	06308	56778	30474	57277	23425	27092	47759
64031	41740	69680	69373	73674	97914	77989	47280	71804	74587	70563	77813
92357	38870	73784	95662	83923	90790	49474	11901	30322	80254	99608	17019
79945	42580	86605	97758	08206	54199	41327	01170	21745	71318	07978	35440
48030	05125	70866	72154	86385	39490	57482	32921	33795	43155	30432	48384
80016	81500	48061	25583	74101	87573	01556	89184	64830	16779	35724	82103
34265	65728	89776	04006	06089	84076	12445	47416	83620	49151	97420	23689
82534	76335	21108	42302	79496	21054	80132	67719	72662	58360	57384	65406
72055	61146	82780	89411	53131	57879	39099	42715	24830	60045	23250	39847
26999	96294	20431	30114	23035	30380	76272	60343	57573	42492	47962	21439
01628	47335	17893	53176	07436	14799	78197	48601	97557	83918	20530	61565
66322	27390	73834	73494	21527	93579	20949	85666	25102	64733	93872	72693
96239	18521	67354	41883	58939	36222	43935	36272	47817	90227	91434	86453
10497	83617	39176	45062	53903	33862	14903	38996	60027	41702	78139	28598
69712	33438	85908	58620	50646	47857	96024	58568	67614	44370	40276	85964
51375	42451	76889	68096	80657	91046	95340	70209	23825	46031	45306	64476

IMPRIMERIE DU CACHOT
75, ANCIENNE ROUTE
1218 GRAND-SACONNEX